Step 3: Compare Results and State Observations

Model 1 performed extremely poorly and was missing a lot of accuracy scores. It's overall accuracy was only 33% even after data normalization and data augmentation.

I tried many different augmentation methods to try to increase diversity of training, and thus the model's accuracy, but nothing made much of a difference. I also tried increasing the number of samples for underrepresented classes but it only made things worse, so I deleted those steps from the data.

	precision	recall	f1-score	support
buoy	0.00	0.00	0.00	13
cruise_ship	0.00	0.00	0.00	47
ferry_boat	0.00	0.00	0.00	15
freight_boat	0.00	0.00	0.00	4
gondola	0.26	0.71	0.38	34
inflatable_boat	0.00	0.00	0.00	2
kayak	0.38	0.11	0.18	44
paper_boat	0.00	0.00	0.00	3
sailboat	0.37	0.67	0.47	70
accuracy			0.33	232
macro avg	0.11	0.17	0.11	232
weighted avg	0.22	0.33	0.23	232

Model 2 performed much better than Model 1, with double the accuracy score of 66%. Adding the MobileNetV2 layer (along with a 70/30 split for train/test) brought great improvements across all categories of boats, represented by all macro and weighted average scores being higher.

Cruise_ship, Kayak, and Sailboat classes had dramatic improvement in the second model.

	precision	recall	f1-score	support
buoy	0.23	0.64	0.33	14
cruise_ship	0.78	0.82	0.80	60
ferry_boat	0.35	0.39	0.37	18
freight_boat	0.05	0.11	0.07	9
gondola	0.71	0.67	0.69	55
inflatable_boat	0.00	0.00	0.00	4
kayak	0.81	0.72	0.76	69
paper_boat	0.67	0.25	0.36	8
sailboat	0.84	0.66	0.74	111
accuracy			0.66	348
macro avg	0.49	0.47	0.46	348
weighted avg	0.72	0.66	0.68	348

Both models struggled with inflatable_boat and freight_boat which means we would need more examples of these classes in the data to better train on them.

Precision and Recall values also increased in model 2 over model 1. This means model 2 was making more accurate predictions and was better at identifying more instances of those classes.

Ultimately, bringing in the pre-trained MobileNetV2 model drastically improved our model's performance and accuracy. 66% accuracy is still relatively low so we would want to increase the size of our data and try to improve the model in other ways like data augmentation and fine-tuning before feeling confident to use it in a professional setting.